DOCUMENT RESUME

ED 434 909 TM 030 138

AUTHOR Schau, Candace; Mattern, Nancy; Zeilik, Michael; Teague, Kathleen W.

TITLE Select-and-Fill-in Concept Map Scores as a Measure of Undergraduate Students' Connected Understanding of Introductory Astronomy.

SPONS AGENCY National Science Foundation, Arlington, VA.
PUB DATE 1999-04-00
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
CONTRACT DUE-9253983
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Astronomy; Comprehension; *Concept Mapping; Diversity (Student); Educational Assessment; Higher Education; Introductory Courses; *Knowledge Level; *Scores; *Undergraduate Students; *Validity

ABSTRACT

This study was designed to explore the validity of scores from a select-and-fill-in (SAFI) concept map assessment as a measure of 130 undergraduate students' connected understanding of introductory astronomy. Scores from SAFI maps created for this purpose possessed high internal consistency and showed large mean increases from the beginning to the end of the astronomy course. The SAFI map scores exhibited large relationships with scores from direct-instruction multiple-choice examinations and with scores from a relatedness ratings measure, taken together and separately, for students overall and when examined separately by gender. Results provide initial evidence of the validity of scores from SAFI maps as measures of connected understanding of introductory astronomy in ethnically diverse undergraduate students. (Contains 26 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 434 909

Select-and-Fill-In Concept Map Scores as a Measure of Undergraduate Students' Connected Understanding of Introductory Astronomy

Candace Schau, Nancy Mattern, Michael Zeilik, and Kathleen W. Teague

University of New Mexico

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Candace Schau

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the 1999 annual meeting of the American Educational Research Association held in Montreal, Canada. For additional copies of the paper or comments about it, contact Candace Schau, Professor; Educational Psychology Program; College of Education; University of New Mexico; Albuquerque, NM 87131; 505-277-3208; cschau@unm.edu. This work was supported in part by the National Science Foundation Grant No. DUE-9253983 to the third author. We would like to thank the students and instructors who participated in our study.

TM030138

BEST COPY AVAILABLE

Abstract

This study was designed to begin to explore the validity of scores from a select-and-fill-in (SAFI) concept map assessment as a measure of undergraduate students' connected understanding of introductory astronomy. Scores from SAFI maps created for this purpose possessed high internal consistency and showed a large mean increase from the beginning to the end of the astronomy course. The SAFI map scores exhibited large relationships with scores from direct-instruction MC exams and with scores from a relatedness ratings (RR) measure, taken together and separately, for students overall and when examined separately by gender. Our results provide initial evidence of the validity of scores from SAFI maps as measures of connected understanding of introductory astronomy in ethnically-diverse undergraduate students.

Select-and-Fill-In Concept Map Scores as a Measure of Undergraduate Students' Connected Understanding of Introductory Astronomy

In the field of education, some of the most useful models that address cognitive processes suggest that knowledge must be organized in order to be accessible from long term memory (see, for example, Anderson, 1995; Farnham-Diggory, 1992). According to such models, expertise is attained by developing rich, accurate, relevant, and accessible sets of organized knowledge (Marshall, 1995). That is, expertise requires "connected understanding," understanding both concepts and the connections among concepts (Schau & Mattern, 1997).

The current work on national educational standards and goals in the U.S. affirms this approach by emphasizing the importance of connected understanding, especially in science learning. The Benchmarks for Scientific Literacy (American Association for the Advancement of Science, 1993) explicitly emphasizes the importance of "coherence and connectedness" (p.XVI) in science learning, stating that "a central Project 2061 premise is that the useful knowledge people possess is richly interconnected" (p. 315). Similarly, in regard to assessment of this kind of learning, the National Science Education Standards (National Research Council, 1996) indicates that "assessment processes that include all outcomes for student achievement must probe the extent and organization of a student's knowledge" (p. 82).

Assessment Formats

Three general assessment formats are important to our work. These formats include concept maps, relatedness ratings techniques, and multiple-choice exams.

Generated Concept Maps

Concept maps have been used extensively in K-12 classrooms, especially to assess

connected understanding in science. In these maps, concept words (or phrases) often are referred to as nodes and placed in a geometric shape, usually an oval or rectangle. The connecting structures between the nodes are called links and usually are represented by labeled lines or arrows. Thus, a proposition consists of two concepts connected by a labeled link; it can be considered the basic unit in connected understanding (see Jonassen, Beissner, & Yacci, 1993; Shavelson, Lang, & Lewin, 1994). For example, "atmosphere surrounds earth" is a proposition commonly included in middle-school earth science.

Although many approaches can be used with concept maps, the most widely used application has students generate their own maps (e.g., Ruiz-Primo & Shavelson, 1996; Shavelson, et al., 1994). In most applications of this format, students arrange important concepts into a map and connect them with links that they label. Students either generate their own concepts or use concepts that are provided. They must have an understanding of interconnected groups of propositions to draw an adequate map, and they can represent their connected understanding of a domain with many different maps. Students see a representation of the domain being assessed take form as they draw their maps. Once students have drawn maps, these maps can be scored quantitatively based on their characteristics using various scoring schemes (e.g., Liu, 1994; Novak & Musonda, 1991; Rice, Ryan, & Samson, 1998).

Student-generated concept maps have at least four limitations as an efficient format for assessing learning. First, students must learn how to draw concept maps and then actually draw them, processes that are time-consuming and can be tedious and frustrating. Some students (and instructors) do not like to and so will not draw concept maps (e.g., Anderson & Huang, 1989; Barenholz & Tamir, 1992). Second, there is no universally accepted and simple scoring system

for generated concept maps (see Shavelson et. al., 1994). Third, although possible to score by computer, this task is not easy. Fourth, the quality of student-generated maps depends heavily on the individual's communication skills (Schau, Mattern, Weber, Minnick, & Witt, 1997).

Relatedness Ratings Techniques

Relatedness ratings assessments were developed and continue to be used primarily by experimental psychologists to assess connected understanding, called structural knowledge in this research area (e.g., Goldsmith, Johnson, & Acton, 1991; Johnson, Goldsmith, & Teague, 1995). This assessment process first involves eliciting students' structural knowledge of a domain indirectly. Each student's structural knowledge is then typically compared with an expert-derived domain structure and can be represented both visually as a map and numerically as a score. The visual characterization requires the use of a computer software program such as Pathfinder (Schvaneveldt, 1990) or a statistical package that includes multidimensional scaling techniques.

For example, a typical relatedness rating (RR) item asks students to indicate the degree of relatedness between a pair of concepts, such as "earth" and "atmosphere", using a Likert-type numerical scale. Students do not characterize the nature of this relationship with a label. Each RR item can be considered one isolated proposition with an unlabeled link. The complete RR measure, then, usually consists of a set of propositions that together implicitly represents the domain being assessed.

The RR format has at least five limitations as an efficient method to assess connected understanding in school settings. First, it is not obvious to teachers or students how RR techniques can assess something as important as connected understanding; thus motivation to

complete these formats can be low. Second, students do not see the map representation of their connected understanding as they complete RR tasks. Third, when maps are generated from RR data, the links in the maps are unlabeled; many researchers and educators believe that labeled links are an important aspect of teaching and learning for connected understanding (e.g., Novak & Gowin, 1984). Fourth, as students rate the relatedness of a specific pair of concepts, they may not perceive the domain context surrounding the pair; judging the degree of relatedness without a sense of the context is difficult. Fifth, most of the published research using RR approaches has been done with college students; it is not clear that elementary or even middle-school students could or would do these kinds of tasks.

Multiple-choice Tests

In research examining the validity of scores from generated concept maps and relatedness ratings measures, multiple-choice (MC) test scores serve as a common comparison measure. Multiple-choice tests may cover knowledge obtained from recent direct instruction (e.g., items in a unit test or a course final) or overall proficiency in one or more broad domains as is assessed by standardized achievement measures.

Although other approaches are possible, it is often the case that a multiple-choice item contains one or a few propositions. Students select a response that often corresponds to a node in one of those propositions. Each item is relatively limited. The complete multiple-choice exam may consist of a set of propositions that together implicitly represents the domain being assessed.

Correlations between student-generated concept map scores calculated using a variety of scoring methods and direct-instruction MC test scores usually fall in the .40 to .60 range (Rice, et al., 1998; Ruiz-Primo, Schultz, Li, & Shavelson, 1998; Ruiz-Primo, Shavelson, & Schultz,

1997). Correlations between map scores and scores from standardized MC achievement measures assessing the same general domain often range between .65 and .85 (e.g., Anderson & Huang, 1989; Rice, et al., 1998).

In the one study we found that examined the relationship between scores from RR measures and from direct-instruction MC scores assessing the same knowledge domain, resulting correlations were .22 and .39 (Diekhoff, 1983). Correlations of scores from RR tasks and from standardized MC measures usually are in the range of .45 - .60 (e.g., Goldsmith, et. al., 1991; Johnson, et. al., 1995).

Fill-in Concept Maps

Another, rarely explored, concept map format uses fill-in maps. Keeping an expert-drawn concept map structure intact, some or all of the concept words and/or linking words are omitted. Students fill in these blanks either by generating the words to use (called "generate-and-fill-in") or by selecting them from a set which may or may not include distractors (called "select-and-fill-in" or SAFI) (Schau, Mattern, Weber, Minnick, & Witt, 1997). Surber (1984) may have been the first to use a fill-in concept map format as an assessment approach. Naveh-Benjamin, Lin, and McKeachie (1995) also used a fill-in approach with hierarchical maps containing unlabeled links.

Each SAFI map item is part of one or usually several explicitly connected propositions. The response selected by the student can be either a node or a link, depending on the kind of SAFI map being used. The map characterizes the domain explicitly, and students explicitly see their additions to the domain representation as they complete the assessment task.

The results from our earlier study (Schau, Mattern, Weber, Minnick, & Witt, 1997)

provided initial evidence that the SAFI concept map format can be used to develop a measure whose scores assess connected understanding in rural, culturally diverse middle-school science students. After initial testing of 22 different map assessment formats, we selected the SAFI map format with less than half of non-consecutive nodes removed from the map and listed on the page in a selection set for further testing. Work with individual students suggested that strategies requiring connected understanding were used to successfully complete SAFI maps based on this format. Item analyses indicated that map scores possessed high internal consistency. If the SAFI map measure assessed science achievement, the eighth grade mean map score should have been and was higher than the seventh grade mean score. If the map measure assessed science achievement for seventh and eighth grade male and female students from Hispanic American, Native American, and White American cultural groups, the map and MC score distributions should have correlated and did correlate strongly for each of these groups. The sizes of these coefficients fell in the middle of the range of values found by researchers exploring the relationships between scores from generated concept map and standardized MC achievement measures (e.g., Anderson & Huang, 1989; Rice, et al., 1998). Clearly these two measures assessed some of the same aspects of science achievement.

In addition to our work, one other study has examined the relationships between scores from SAFI concept maps and from other measures. Map score correlations with high school students' scores from a direct-instruction MC test were .37 (for maps requiring students to fill in a selection of missing nodes) and .65 (for maps requiring students to fill in a selection of missing links). The authors reported that the lower relationship occurred because the SAFI node maps were too easy for the students, causing a ceiling effect (Ruiz-Primo, et al., 1998). Their study did

not include a standardized MC domain measure.

It appears that the SAFI concept map format has the potential to overcome at least some of the limitations associated with generated concept maps, RR formats, and MC items. However, a potential major disadvantage to SAFI maps is that students are provided with a representation of the domain structure being assessed. Students do not either create a unique structure representing their own connected understanding (as occurs when students generate concept maps) nor are their responses used to generate a representation of their connected understanding of the domain (as can occur with data from relatedness ratings techniques). Until our work, there was little evidence to indicate that fill-in map scores assess students' connected understanding.

Purpose

This study was designed to begin to explore the validity of SAFI concept map scores as a measure of culturally diverse post-secondary students' connected understanding of introductory astronomy. It included the development of a set of SAFI maps for use with undergraduate students enrolled in an introductory astronomy course and an evaluation of the convergent validity of these map scores.

We hypothesized that 1) a SAFI map measure assessing connected understanding of introductory astronomy would show good internal consistency. We also hypothesized that SAFI map scores would assess students' connected understanding of introductory astronomy. Therefore, these scores should 2) show a large mean gain from the beginning to the end of the course, and 3) relate positively with at least medium effect sizes to MC direct-instruction exam scores and to RR scores for students overall and grouped by gender. Unfortunately, our sample did not include enough students to examine these relationships by self-reported ethnic/cultural

group. We also examined the relationships of the score distributions from RR and MC measures (implicit measures of connected understanding) together and uniquely with the scores from SAFI maps (an explicit measure of connected understanding).

This study was an integral part of instructional evaluation; data collection took place during regular class periods. As such, it could not involve tight experimental controls nor could all desired measures be administered to the students. For example, it was not possible to administer measures to address divergent validity concerns in the traditional manner; instructional time was too limited to include measures that were hypothesized to be unrelated to achievement.

General Analysis Information

SPSS 7.5 for Windows and 6.1 for the Mac were used for all statistical analyses. When statistical test results are reported, effect sizes also are given. Cohen's (1988) suggested effect size measures and guidelines were used to evaluate the practical importance associated with each result. For mean differences, d was calculated. d values of about .2 or less were considered small; about .5, medium; and about .8 or more, large. For correlations, the Pearson Product Moment correlation coefficient itself served as the effect size. Values around .1 or less were considered small; around .3, medium; and around .5 or more, large. For multiple regression results, f^2 served as the effect size. Values around .02 or less were considered small; around .15, medium; and around .35 or more, large.

Method

Participants and Setting

Data were collected in one of the three sections of the undergraduate introductory

astronomy course for students who were not majoring in science offered at the largest state-supported university in New Mexico. The section met twice per week for 75 minutes per meeting. The instructor (the third author) had designed instruction in the course to be conceptually-based. In addition to the traditional delivery methods of lectures, demonstrations, and computer simulations, he also used instructional concept maps and small student-centered focused discussion groups.

Of the originally enrolled 163 students, 130 completed the course. Fifty-eight percent reported that they were female. Sixty-three percent reported that they were White American, 20% Hispanic American, and the remainder African American, Native American, or "other".

Measures

A SAFI concept map measure and a RR criterion measure were developed. These measures were designed to possess global relevance to the domain of introductory astronomy for non-science majors and to be appropriate for use in this specific course. The content included in these measures was developed using the best procedures possible, given the applied setting of the course. Both measures were pilot tested in this same course during previous semesters and revised for use in this study. No attempt was made to include the same concepts in these achievement measures that were tested in the MC course exams, the third measure used, although there was overlap since these concepts were the important ones in the course.

SAFI Concept Map Measure

A set of three novel master concept maps was created, each representing global connected understanding of major concepts covered in the course. Concepts were enclosed in ovals and connections between concepts were represented by labeled directional arrows. These maps were

not used in instruction nor were they available to students except during their pre- and post-administration. Thirty non-sequential ovals (10 per map or 35% of the 85 total map nodes) were left blank; all links were left intact. The concepts missing from these blank ovals were listed in a selection set located in a corner of each map. Students selected a response from the list of 10 and wrote it into the corresponding blank node; responses could be used more than once. Students then transferred the response letter associated with each answer to a scannable answer sheet and continued on to the next map. The SAFI map measure was scored based on the percentage of correct responses to 28 blank nodes. Two blank nodes were eliminated from the scoring, one because it was used to identify non-participants and the other because it was used as an example. See Figure 1 for one of these maps.

Insert Figure 1 about here

Relatedness Ratings Measure

Three tasks were involved in developing the RR measure: 1) selecting essential concepts, 2) determining the associations among selected concept pairs to identify related and unrelated pairs, and 3) creating the student measure. To identify the essential concepts, the course instructor first compiled a list of 200 of the more global concepts found in the glossaries of the two undergraduate introductory astronomy texts that he had authored. This list of concepts was sent to 30 experienced post-secondary astronomy instructors from across the US; 18 returned responses. Using a five-point Likert scale, each instructor was asked to rate the importance of students' understanding of the concept by the end of a one-semester introductory astronomy

course for non-science majors. The 120 concepts with the highest mean importance scores and the lowest standard deviations (indicating a high level of agreement that these concepts were important) were selected.

In the associated pairs task, these concepts were randomly split into two sets of 60 target concepts each. Half of the same 30 instructors was sent the first set of target concepts while the other half received the second set. The instructors were asked to select from the entire remaining list of 119 concepts those that were highly related to each of the target concepts. Eleven of the 15 who received the first set of target concepts returned their responses while only six who received the second set returned theirs. To quantify the generality/specificity of a concept in the domain of introductory astronomy, each target concept was given a "scope" score which was the number of times any of the 17 experts selected that concept as associated with any other concept. To quantify the relatedness of pairs of concepts, each pair was given an "association" score which was the number of times across experts those two concepts were associated as a pair. Because of the disparate return rates between the two sets of target concepts, only concepts from the first set were used for determining association scores. In creating the RR measure, related pairs were selected based on high association scores for the pair and high scope scores for both concepts in the pair, as well as appropriateness for this specific course as determined by the course instructor. Unrelated pairs used in the RR measure were selected based on low association scores, high scope scores, and appropriateness for this course. This process yielded 95 pairs, 58 related and 37 unrelated pairs.

The pairs were randomly assembled into the students' RR measure. Students judged the relatedness of each pair on a five-point Likert scale (1 was Unrelated, 3 was Moderately Related,

5 was Highly Related). Students' individual scores were computed as the point-biserial correlation between their responses to each pair and each pair's designation by the experts as related (1) or unrelated (0). For more information about the general technique, see Johnson, et al. (1995).

Course Exam Measure

The course instructor developed four MC exams over the course of the semester; the first three contained 50 items while the last included 30 items. Scores on each exam were based on the percentage of correct responses. Individuals' three highest exam scores were averaged for use in the analyses.

Procedure

Students completed the SAFI concept map measure and the RR measure twice, once at the beginning and again at the end of the semester. Due to the time constraints imposed by the length of the class period and concerns about participant burn-out, the students did not complete both measures during class; the RR measure was always completed outside of class because it was unlikely that students would seek out information that could invalidate their scores. During the first week of the semester, students completed the map measure during part of one class period and the RR measure outside of class. They again completed the map measure during class after they had finished their final exam; the RR measure was again completed on their own time and returned at the beginning of the final exam period. Students were volunteers who could stop participation at any time. Students who completed all SAFI map and RR measures received credit for two extra homework assignments for their participation. They completed their exams during class.

Data Screening

In order to be included as a participant on the SAFI concept map measure, students had to select a response for at least 75% of the nodes on each of the three maps and select the correct answer for one obvious node. For the pre-course administration, 17 students were non-participants; for the post-course administration, 10 students were non-participants. Participants on the RR measure were defined as those students who selected a response to at least 75% of the pairs. Two students were non-participants on the pre-course administration; one on the post-course administration. Students had to complete at least three exams to be considered participants; four students were non-participants on the exam measure.

Scores that fell over 3 standard deviations from their cell means and were discontinuous from their closest neighboring scores were considered outliers. After examination of these subjects' scores, and following the reasoning of Tabachnick and Fidell (1996) as well as others, they were eliminated and their cells checked again until no more outliers were identified. Two students were low outliers on the exam measure; no outliers occurred on either the pre- or the post-course administrations of the SAFI map or the RR measures.

The number of students involved in each analysis varied depending on the number of students who completed the corresponding measures, as well as the number who were eliminated as non-participants and outliers. From the pre-course administration, the analysis samples for the SAFI map measure consisted of 133 students; for the RR measure, 138 students. From the post-course administration, the analysis samples for the SAFI map measure contained 93 students; for the RR measure, 118 students. The exam measure analysis sample contained 124 students. Pair-wise elimination was used in all correlational analyses.

Results

We examined each measure for internal consistency, changes in mean scores from the SAFI map and the RR measures across the semester, and correlations between SAFI map score distributions and score distributions from the RR measure and the exam measure at the end of the course.

Internal consistency

Item analysis procedures were employed to examine the internal consistency of the SAFI map scores obtained from the post-course administration. Supporting Hypothesis 1, the value of Cronbach's alpha was .83 (N=93). All items functioned adequately; the alpha value could not have been increased by more than .01 through item elimination. The split-half reliability of the RR measure was .66 (N=118); for this analysis, each subject was given two scores, with each score containing randomly-selected equal numbers of related and unrelated item pairs. The pair-wise correlations between the four exams varied from .36 to .61, with an average intercorrelation of .52.

Score Relationships

Supporting Hypothesis 2, SAFI map scores showed a mean increase across the semester from 30% correct (SD = 11%) at the beginning of the course to 50% correct (SD = 19%) at the end of the course, dependent $t(83) = 10.09$, $p<.0005$, $d = 1.30$. Mean RR scores also increased from .14 (SD = .14) to .38 (SD = .20), dependent $t(102) = 13.17$, $p<.0005$, $d = 1.35$. Both increases were large.

Supporting Hypothesis 3, correlations between post-course SAFI map scores and each additional measure were statistically significant and large in size for students overall and for

female and male students. In addition, correlations between post-course RR scores and MC exam scores were statistically significant and medium or large in size for students overall and for students grouped by gender. See Table 1.

Insert Table 1 about here

Using pair-wise deletion of missing data in multiple regression, direct MC and RR scores were statistically significantly related to SAFI map scores, $F(2,86) = 26.79$, $MSE = 19.73$, $p < .0005$, $R^2 = .38$, $f^2 = .62$, a large effect size. Uniquely, RR scores were positively and statistically significantly related to SAFI map scores, $t(86) = 4.13$, $p < .0005$, $sr = .35$, $pr = .41$, $f^2 = .20$. Uniquely, MC scores also were positively and statistically significantly related to SAFI map scores, $t(86) = 3.91$, $p < .0005$, $sr = .33$, $pr = .40$, $f^2 = .18$. The effect sizes of both unique relationships were medium.

Discussion

These findings provide evidence that the SAFI concept map format can be used to develop a measure whose scores assess connected understanding in culturally diverse undergraduate introductory astronomy students. All hypotheses were supported. Scores from the SAFI map measure showed good internal consistency. Both SAFI map scores and RR scores showed approximately equal (in standard deviation units) large increases across the semester.

The consistently large relationships between direct-instruction MC exam scores and SAFI map scores suggest that the map scores assessed some of the same kinds of course knowledge as those assessed by MC course exam scores. The sizes of these coefficients fell within the range of

correlations reported in research exploring the relationships between generated concept maps and direct-instruction MC scores (Rice, et al., 1998; Ruiz-Primo, Schultz, Li, & Shavelson, 1998; Ruiz-Primo, Shavelson, & Schultz, 1997). Similarly, the consistently large relationships between RR and SAFI map scores suggest that map scores also assessed some of the same kinds of structural course knowledge as those assessed by RR scores. However, there was a sizeable amount of variation in SAFI maps scores that was not shared with MC and RR scores, taken either together or separately.

The SAFI map format overcomes the limitations described for generated concept maps. First, students learn to complete fill-in maps quickly, and many like completing them. Second, although other scoring systems are possible, a simple accepted scoring system exists; the fill-in map responses can be scored as correct or incorrect. Third, when using the right-wrong scoring system, fill-in map responses can be scored easily by computer. Fourth, students with lower levels of communication skills can complete SAFI maps.

In addition, the SAFI map format also overcomes the limitations associated with RR techniques. First, students and teachers believe that the maps assess connected understanding. Second, students see a visual representation of the domain as they complete their maps. Third, the maps contain labeled links; in fact, students must use the links (as well as the nodes) to complete their maps. Fourth, each map provides an explicit context to students. Fifth, SAFI maps can be used easily with students beginning at least in seventh grade, and we believe that much younger students could complete maps created using this format.

Assessment formats based on concept maps, relatedness ratings (or any other kind of structural knowledge technique), and multiple-choice exams (or any other kind of traditional

classroom assessment) rarely are included together in the same research studies. The dearth of this kind of comparative study makes it difficult to sort out what aspects of connected understanding these formats do and do not assess. Our study makes an important beginning contribution in this area.

Each year, millions of students take introductory science classes. An efficient method for assessing students' connected understanding in classroom settings will be an invaluable tool for researchers and educators concerned with their students' cognitive growth. We believe that we have taken initial steps toward achieving this goal. Our evidence indicates that scores from good SAFI concept map measures assess connected understanding in undergraduate introductory astronomy students.

References

Anderson, J. R. (1995). Learning and memory: An integrated approach. New York: Wiley.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

Anderson, T. H., & Huang, S. C. C. (1989). On using concept maps to assess the comprehension effects of reading expository text. (Report No. 483). Champaign, IL: University of Illinois (ERIC Document Reproduction Service No. ED 310 368).

Barenholz, H., & Tamir, P. (1992). A comprehensive use of concept mapping in design instruction and assessment. Research in Science and Technological Education, 10, 39-52.

Cohen, J. (1988). Statistical power analysis for the behavioral science (2nd ed.). Hillsdale, NJ: Erlbaum.

Diekhoff, G. M. (1983). Testing through relationship judgments. Journal of Educational Psychology, 75, 227-233.

Farnham-Diggory, S. (1992). Cognitive processes in education (2nd ed.). New York: HarperCollins.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. Journal of Educational Psychology, 83, 88-96.

Johnson, P. J., Goldsmith, T. E., & Teague, K. W. (1995). Similarity, structure, and knowledge: A representational approach to assessment. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), Cognitively diagnostic assessment (pp. 221-249). Hillsdale, NJ: Erlbaum.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). Structural knowledge: Techniques

for representing, conveying, and acquiring structural knowledge. Hillsdale, NJ: Erlbaum.

Liu, X. (1994, April). The validity and reliability of concept mapping as an alternative science assessment when item response theory is used for scoring. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Marshall, S. (1995). Schemas in problem solving. Cambridge: Cambridge University Press.

National Research Council (1996). National science education standards. Washington: National Academy Press.

Naveh-Benjamin, M., Lin, Y, & McKeachie, W. J. (1995). Inferring students' cognitive structures and their development using the "Fill-in-the-Structure" (FITS) technique. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), Cognitively diagnostic assessment (pp. 279-304). Hillsdale, NJ: Erlbaum.

Novak, J. D., & Gowin, D. B. (1984). Learning how to learn. New York: Cambridge University Press.

Novak, J. D., & Musonda, D. (1991). A twelve year longitudinal study of science concept learning. American Educational Research Journal, 28, 117-153.

Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? Journal of Research in Science Teaching, 35, 1103-1127.

Ruiz-Primo, M. A., Schultz, S. E., Li. M., & Shavelson, R. J. (1998). Comparison of the reliability and validity of scores from two concept-mapping techniques. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. Journal of Research in Science Teaching, 33, 569-600.

Ruiz-Primo, M. A., Shavelson, R. J., & Schultz, S. E. (1997). On the validity of concept map-based assessment interpretations: An experiment testing the assumption of hierarchical concept maps in science. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Schau, C., & Mattern, N. (1997). Use of map techniques in teaching applied statistics courses. The American Statistician, 51, 171-175.

Schau, C., Mattern, N., Weber, R. W., Minnick, K., & Witt, C. (1997, April). Use of fill-in concept maps to assess middle school students' connected understanding of science. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Schvaneveldt, R. W. (Ed.) (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

Shavelson, R. J., Lang, H., & Lewin, B. (1994). On concept maps as potential "authentic" assessments in science (CSE Tech. Rep. No. 388). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Surber, J. R. (1984). Mapping as a testing and diagnostic device. In C. D. Holley & D. F. Dansereau, D. F. (Eds.). Spatial learning strategies: Techniques, applications, and related issues (pp. 213-233). Orlando: Academic Press.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: HarperCollins.

Table 1

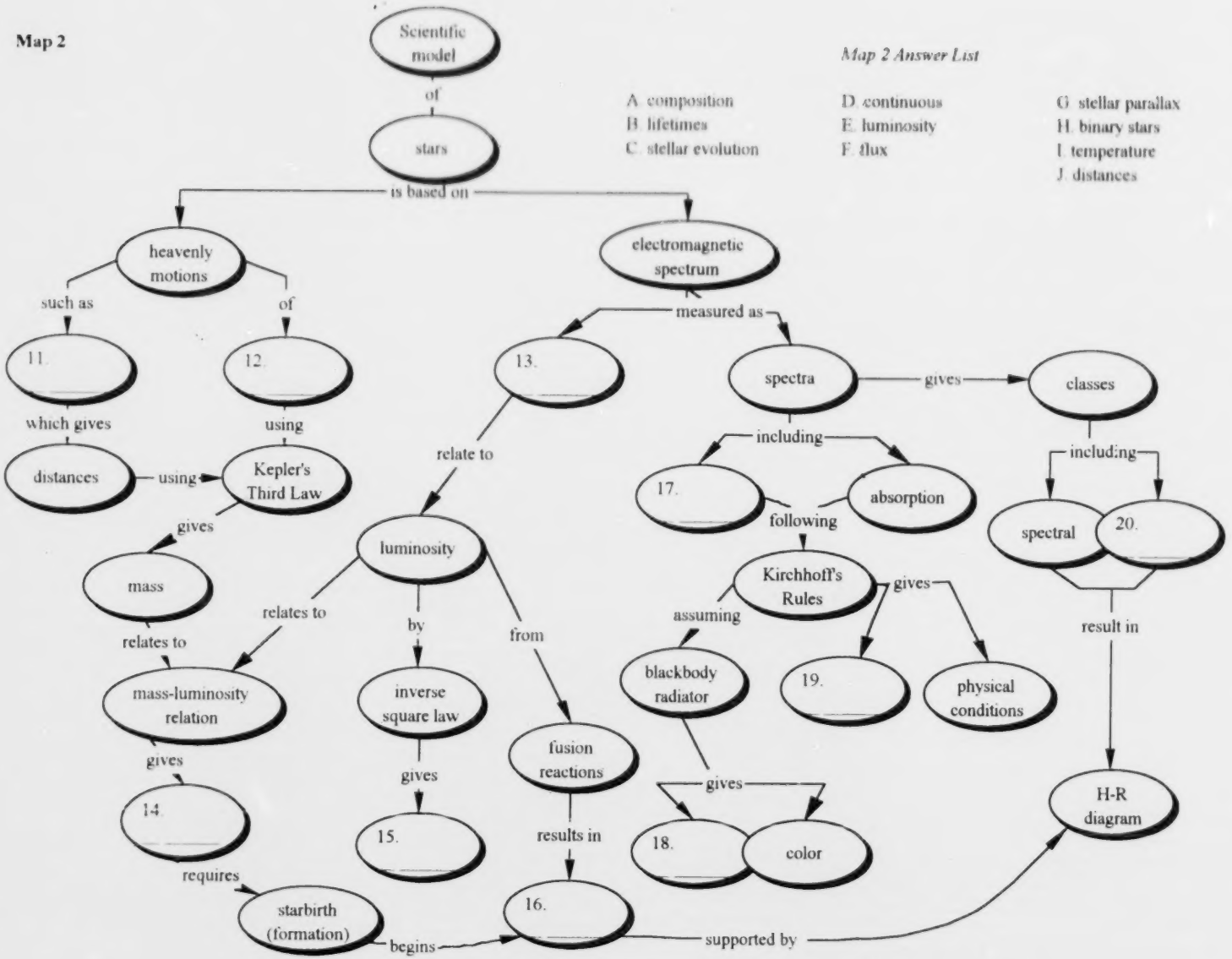
Correlation Results for Introductory Astronomy Students

	Scores		
Group	Maps & MC Exams	Maps & RR	RR & MC Exams
Overall			
r	.51	.52	.40
n	93	89	115
p	<.0005	<.0005	<.0005
interpretation	large	large	medium/large
Male			
r	.47	.48	.46
n	39	37	51
p	.003	.003	.001
interpretation	large	large	large
Female			
r	.53	.55	.33
n	54	52	64
p	<.0005	<.0005	.007
interpretation	large	large	medium

.Figure Caption

Figure 1. One of three select-and-fill-in concept maps designed to yield scores that assess undergraduate students' connected understanding of introductory astronomy

Map 2



Map 2 Answer List

A. composition
B. lifetimes
C. stellar evolution
D. continuous
E. luminosity
F. flux
G. stellar parallax
H. binary stars
I. temperature
J. distances

BEST COPY AVAILABLE